DOCUMENT RESUME

ED 076 197 LI 004 295

AUTHOR	Line, Maurice B.; And Others
TITLE	Patterns of Citations to Articles within Journals: A Preliminary Test of Scatter, Concentration and Obsolescence.
INSTITUTION	Bath Univ. of Technology (England). Univ. Library.
PUB DATE	Oct 72
NOTE	37p.; (15 References)
AVAILABLE FROM	Bath University Library, Claverton Down, Bath BA2 7AY, England (50 pence)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Bibliographic Citations; Information Utilization; *Library Material Selection; *Obsolescence; *Periodicals; *Technical Reports
IDENTIFIERS	*Scientific Periodicals

ABSTRACT

If citations are concentrated on a relatively small number of articles within individual journals as well as between different journals, selections of key articles can be published that should satisfy a high percentage of library demand for back runs at low purchase cost. To test this hypothesis, a test was carried out on self-citations in three journals in different scientific disciplines. The results show a considerable degree of concentration, though it differs from journal to journal. The relationship of self-citations as indicators of use is discussed, and a comparison was made with citations to the same journals in "Science Citation Index." The papers that were most frequently cited were different at different periods. The data also provided an opportunity to compare synchronous and diachronous obsolescence. Their differences are discussed and the latter is shown to depend on more variables than the former, but differences between the two were not demonstrable. Obsolescence rates were measured as the annual rate of decay in the probability that any one paper of a given age would be cited in any one paper in the same journal of a given year. (Author/NH)

Bath University Library

PATTERNS OF CITATIONS TO ARTICLES WITHIN JOURNALS: A PRELIMINARY TEST OF SCATTER, CONCENTRATION AND OBSOLESCENCE

Maurice B. Line (National Central Library)
and
A. Sandison (National Reference Library for Science & Invention)
assisted by
Jean MacGregor (Bath University Library)

October 1972

PATTERNS OF CITATIONS WITHIN JOURNALS

Responsibility for Parts 1 and 2 rests with Mr Line and Miss MacGregor, for Part 3 with Mr Sandison.

CONTENTS

Figures:

ABSTRACT

If citations are concentrated on a relatively small number of articles within individual journals as well as between different journals, selections of key articles can be published that should satisfy a high percentage of library demand for back runs at low purchase cost. To test this hypothesis, a test was carried out on self-citations in three journals in different scientific disciplines (bacteriology, applied chemistry and mathematics). The results show a considerable degree of concentration, though it differs from journal to journal. The relationship of self-citations to citations as indicators of use is discussed, and a comparison was made with citations to the same journals in Science Citation Index. The papers that were most frequently cited were different at different periods.

The data also provided an opportunity to compare synchronous and diachronous obsolescence. Their differences are discussed and the latter is shown to depend on more variables than the former, but differences between the two were not demonstrable. Obsolescence rates were measured as the annual rate of decay in the probability that any one paper of a given age would be cited in any one paper in the same journal of a given year. There was no evidence that synchronous rates were altering over the 11 years 1960-70, but the rates were lower for older papers than for younger, so that the decay was not strictly exponential. Rates differed significantly for younger papers between the three journals, as did the average numbers of within-journal citations per paper.

1. INTRODUCTION

1.1 Citation concentration and its potential practical value

It is now well established by a number of studies that a relatively small proportion of journals in a subject field accounts for a high percentage of use at any given time, whether measured by library consultation and borrowing or by citations. The value of these studies is that they can provide some guidance to libraries as to the number of periodicals they need to buy to satisfy a reasonable percentage of demand, and as to the actual titles they might purchase (though this will of course depend very much on local interests). If sufficient data can be collected, it may help a librarian to estimate the level at which, and identify the titles for which, he should cease to purchase journals and should instead depend on inter-library loan.

Some examples of this general law are given in Table 1.

Table 1

Examples of citation concentration

Subject field	% of journals account for	% of articles cited
Developmental endocrinology:[6]	14	75
Non-IEEE journals cited in IEEE publications:[2]	10	75
IEEE publications cited in IEEE publications:[2]	13 24	75 90
Chemical engineering:[1]	17	75
Mechanical engineering:[1]	17	75
Metallurgical engineering:[1]	12	75
Journals cited in Physical Review:[5]	2	84
Geology:[4]	30	75

Apart from the last two, these figures are remarkably similar. It should be noted that these figures all underestimate the concentration of citations in a limited proportion of journals, since the studies are concerned

only with journals cited, not all journals which might be cited (indeed, it would be impossible to establish with any precision the total number of journals of potential relevance to a subject).

This limitation does not apply to much the largest citation study, that by Martyn & Gilchrist[8] of citations of UK scientific and technological journals found in Science Citation Index; they were able to relate these to the total population of relevant UK journals, as well as to all titles cited. In this study, 6 per cent of all titles (19 per cent of all cited titles) accounted for 90 per cent of all citations, and 9 per cent of all titles (28 per cent of all cited titles) accounted for 95 per cent of all citations. (It may be noted that a rather similar distribution has been found with loans from the NLLST). It will be noted that the two ratios - of citations to journals cited, and to all journals - are very different, the latter being not dissimilar to, though still lower than, most of the examples quoted above. In the present paper, the problem of defining disciplinary boundaries and establishing total populations of articles does not arise, since the total population of citable articles must be equal to the number of articles in each journal.

Although it can be taken as established that a relatively few journals account for a high proportion of use, it is by no means self-evident that within journals a similar pattern would apply. Indeed, it could well be supposed that 'core' journals would attract to themselves a concentration of high quality articles, which would not vary very greatly in the number of citations they received, although there would obviously be some articles of quite special significance which became classics. Rather than the 9 per cent:95 per cent pattern found in the Martyn and Gilchrist study, one might perhaps expect to find a 30 per cent:70 per cent pattern. Even a pattern such as this would mean that there was within that journal a core of more heavily cited articles, and if these could be made available separately from the main journal, they could presumably be published at a lower cost than the full journal, and enable librarians to satisfy a higher percentage of demand for the same price (or the same percentage of demand for a lower price). This could not of course be done currently, since while it may be known which journals are likely to be most heavily cited in future (from their past citation ranking), it would hardly be possible to estimate exactly which articles will be more heavily cited than others

in future. However, once the heavily cited articles were identified, they could be reproduced in a much smaller number of volumes than the original journal; for example, if a journal which produced 1,000 articles in the decade 1951-60 were shown to have 100 especially significant articles, accounting for 80 per cent of citations to the journal, these 100 articles could be reproduced in one volume for the whole decade.

A particularly attractive feature of this idea is that, unlike many ideas suggested and favoured by librarians for reducing their costs, it would benefit publishers as well as librarians. The number of back sets a publisher has available after a number of years is very small indeed, or non-existent; he aims to make a profit, or keep going, on the sale of current issues, not of back volumes. If he could produce selections of articles on the lines suggested, these would be a saleable commodity and provide an added source of profit. He could reprint (e.g. by offset litho) at fairly low cost. The benefits to librarians are obvious; apart from the savings in cost from their not having to buy complete back runs of serials or rely heavily on inter-library borrowing, a great deal of space would be saved. An alternative would be the provision of xerox copies of individual articles, but these are hard to store and make available through catalogues; a bound volume covering a few years represents a better solution. The second-hand periodical dealer would not welcome the idea, since he would no longer be able to sell back runs so readily; there would always be some libraries ready to buy them, but there would be more libraries ready to discard them.

It is true that collections of reprinted articles already appear from time to time; but nearly all of these are from a variety of journals, and they probably do very little to reduce demand for the original volumes, since the articles in question will normally be cited by reference to the original journals, not to the collections. In any case, it is doubtful if these collections are based on any objective evidence on extent of past use; this is not to decry them, merely to point out that they do not in their present form present an alternative to the suggestion made here.

1.2 Citations and self-citations* as indicators of journal use

Citations are of course an imperfect index of use; people do not cite all they use, even if they write up their research in articles, and for that matter they may not read all they cite. In particular, some journals are widely read for the purpose of keeping up to date on general or technical developments but do not aim to include the kind of 'archival' article that gets cited. Vickery[14] compares citations with four other indicators of the use of periodicals, by ranking the top 89 periodicals obtained according to each method. The differences were very large; for example, 40 of the top 89 journals obtained from Science Citation Index were in medicine, compared with 11 obtained from the World List of Scientific Periodicals (where rankings were according to numbers of holdings in libraries), and compared with 7 obtained from the lists of journals borrowed from the NLLST. In technology, the respective figures are 7 (SCI), 19 (WL), and 51 (NLL).

One would expect citations to be a better guide to the relative value of archival journals than of all journals, for the reasons stated above, though there is no objective evidence, so far as we know, to support this. If this is so, libraries can still use citation studies as a guide to the purchase of back runs of journals (as compared with current issues), since they would not normally consider purchasing back runs of non-archival journals. In the present paper, we are considering only journals which are known to be archival, and for which there is already known to be some demand. The question of the relative value of different journals does not therefore arise; what is at issue is the relative value of articles within given journals, and here one would expect citations to be a good guide, since many of the problems associated with their use as indicators do not apply in these circumstances.

To test the hypothesis that a relatively small number of articles within a journal accounts for a high percentage of use, ideally one would take a selection of journals in several subjects - say 30 or 40 - construct a data base of citations to all of these journals from all journals over a

---

* 'Self-citation' is used in this paper for citation to a journal by that journal, not for citation by an author to papers by himself.

number of years - say 10 - and analyse the results. The amount of work involved in this would be immense, although a fair approximation to it could be made by use of Science Citation Index tapes. However, a first test of the hypothesis can be made rather more simply.

Several studies have shown that there are likely to be more citations to articles appearing in a given journal by that journal than by any other single journal; this is not surprising, since a journal is bound to be more like itself than any other journal, however apparently close in subject. Also, authors are in general more likely to cite themselves than other authors are, and an author who is writing articles on the same subject may be likely to try and publish them in the same journal. Finally, readers of a particular journal, as being of central relevance to their own research, would be likely both to cite articles in it and to think of it as a vehicle for publication of their own articles. Whatever the reasons, the pattern is quite clear, particularly with high status journals.

One of the most useful studies for comparing self-citations with all citations to a journal is that by Xhignesse and Osgood[15], who analysed the citations in 21 psychology journals in 1950 and 1960, and constructed from them a matrix of citations to the same 21 journals. This study shows clearly that journals are in general far more likely to cite themselves than most other journals are, and that, within the matrix, self-citations represent a substantial proportion of all citations to a journal. To only 6 of the 21 journals were self-citations not more numerous than citations by any other single journal. Self-citations represented 32 per cent of all 1960 within-matrix citations, and 12 per cent of all 1960 citations by the 21 journals. The percentage of within-matrix citations to each journal represented by self-citations varied from 10 per cent to 52 per cent (the figures for each journal were, in percentages, 10, 12, 21, 24, 24, 25, 30, 30, 31, 32, 32, 32, 33, 37, 38, 39, 40, 43, 45, 50, 52). It must be borne in mind that all the 21 journals analysed were significant ones, and that it is not possible to tell from the data the percentage of total citations to a journal that is represented by self-citations (this would require analysis of a vast number of journals, though Science Citation Index would give a much more complete picture); nor is it possible to tell whether the self-citations are representative of all citations to a journal. Within-matrix citations accounted for 38 per cent of all citations found in those journals. It is very unlikely that within-matrix

citations account for 38 per cent of all citations to those journals, since they are, as mentioned, all 'status' journals, likely in general to be more cited than citing. If within-matrix citations accounted for as much (or as little) as 20 per cent of all citations to those journals, self-citations would range between 2 per cent and 10 per cent of all citations to the journals studied. It would be interesting to have accurate figures on this, for different subjects and for journals of different status.

At any rate, it was thought that an analysis of citations within a journal would probably offer a reasonable preliminary test of our hypothesis. If no pattern appeared during analysis, it would be unlikely that any significant patterns would appear from analysis of citations appearing in a much larger number of journals. However, as explained later, a further check was made in Science Citation Index, for comparative purposes.

It is possible that if citations from other journals to that journal were taken into account, the pattern revealed would be different; there may, for example, be some articles which are on the fringe of the subject covered by the journal, and hence are more likely to be cited outside it than within it. Conversely, some articles central to the subject may be heavily cited within the journal, but hardly cited at all outside it. If both these factors apply, they would tend to cancel one another out, but in any case the results obtained would not be expected to be more than an approximation.

2. SCATTER AND CONCENTRATION

2.1 Choice of journals for testing

The choice of journals for testing purposes is not easy. Journals which are very heavily cited are not ideal for testing, as libraries would normally expect to buy and retain these journals if at all possible. The Biochemical Journal, for example, probably contains so many heavily cited articles that even if the hypothesis is true, and a small number of articles accounts for a high proportion of citations, even the relatively lowly cited articles would still be of importance, and a library would wish to have a back run of the journal. At the other extreme, many journals are hardly cited at all, and no testing on these is possible. The journals most appropriate for testing are the second level of journals, namely those that are not the central core journals, but that receive sufficient use for many libraries to buy them.

It may well be that the pattern of citation differs significantly from subject to subject, and from journal to journal. Some journals may account for a large number of citations, but these may be to a very few articles, all very heavily cited, or to a large number of articles, all lightly cited. For the purposes of this study, an initial test was made on Journal of Applied Bacteriology, Journal of Applied Chemistry, and Journal of the London Mathematical Society. These were chosen because they represented three different subjects; also because, with respective starting dates of 1938, 1951 and 1926,* the number of citations generated was likely to be large enough to yield valid results without being unmanageable.

2.2 The test and its results

Some articles may be very heavily cited for a limited period, and then drop out of the citation race almost completely; others, on the

---

*In fact, citations to JLMS before 1940 were very few, and not included in the analysis.

other hand may come to be recognised gradually as important articles. Some may have a consistent citation level from year to year. To allow for the possibility of heavy but temporary citation, a period of eleven years was chosen as that to be covered by the source journals; i.e., all volumes from 1960-1970 were checked for citations to articles in the same journal from the beginning up to and including 1970. Cited articles published 1951-59 were analysed separately, to allow a valid comparison to be made between the three journals (which, it will be remembered, all have different starting dates), and to compare within each journal the citation pattern over all volumes and over a period of nine years only.

Tables 2-4 show the results of the analysis of citations to all articles since the beginning of each journal; Figure 1 shows the same results in graphic form. As will be seen, the distributions for each journal are quite different, with a far higher concentration of citations in JLMS and JAC than in JAB; the great majority of articles in JLMS, and the majority of those in JAC, were not cited at all. In fact, the total number of citations in JLMS was small, and no article was cited more than 10 times, whereas one article in JAB was cited 22 times. These differences may merely reflect different citation practices between subjects; it would be premature to conclude that there is a much greater concentration of use in JMLS than in JAB. It would of course require many more examples to show whether the journals are representative of their subjects, though it would perhaps be expected that the 'harder' the subject, the greater the concentration of citations is likely to be, since articles in them are likely to be rather more precise and specific, and hence to be of less wide potential relevance.

Table 2

Journal of Applied Bacteriology:
citations to articles in JAB 1938-70 by articles in JAB 1960-70

	Articles cited			Citations		
No. times cited	N	N (cum)	% (cum)	N	N (cum)	% (cum)
22	1	1	0.1	22	22	1.9
20	1	2	0.2	20	42	3.6
17	1	3	0.2	17	59	5.0
13	1	4	0.3	13	72	6.1
12	1	5	0.4	12	84	7.1
10	2	7	0.6	20	104	8.8
9	2	9	0.7	18	122	10
8	3	12	1.0	24	146	12
7	4	16	1.3	28	174	15
6	9	25	2.1	54	228	19
5	16	41	3.4	80	308	26
4	34	75	6.2	136	444	38
3	58	133	11	174	618	52
2	136	269	24	272	890	75
1	291	560	56	291	1181	100
0	644	1204	100	0	1181	100

Read as follows: (eighth line of figures) 3 articles were cited 8 times, representing 24 citations; 12 articles (1% of all articles published 1938-70) account for 146 citations (12% of all citations in 1960-70 to articles published 1938-70).

Table 3

Journal of Applied Chemistry:
citations to articles in JAC 1951-70 by articles in JAC 1960-70

	Articles cited			Citations		
No. times cited	N	N (cum)	% (cum)	N	N (cum)	% (cum)
14	1	1	0.1	14	14	1.7
12	2	3	0.2	24	38	4.5
10	2	5	0.3	20	58	6.9
9	1	6	0.3	9	67	8.0
8	1	7	0.4	8	75	8.9
7	3	10	0.5	21	96	11
6	4	14	0.7	24	120	14
5	3	17	0.9	15	135	16
4	22	39	2.0	88	223	27
3	48	87	4.5	144	367	48
2	99	187	9.7	198	565	67
1	274	460	24	274	839	100
0	1462	1922	100	0	839	100

Table 4

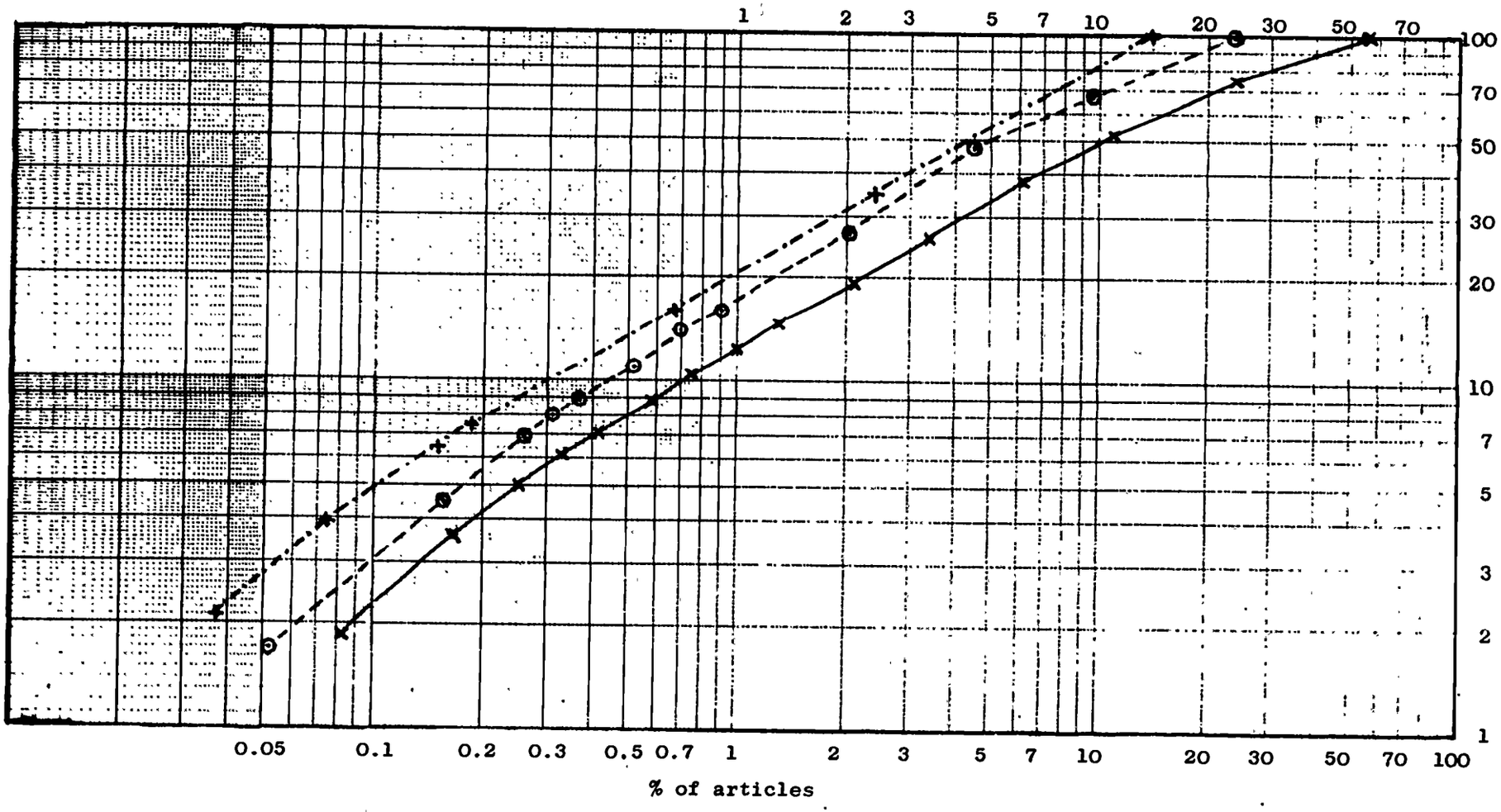
Journal of the London Mathematical Society:
citations to articles in JLMS 1940-70 by articles in JLMS 1960-70

	Articles cited			Citations		
No. times cited	N	N (cum)	% (cum)	N	N (cum)	% (cum)
10	1	1	0.1	10	10	2.1
8	1	2	0.1	8	18	3.9
6	2	4	0.1	12	30	6.4
5	1	5	0.2	5	35	7.5
3	13	18	0.7	39	74	16
2	46	64	2.4	92	166	35
1	302	366	14	302	468	100
0	2309	2675	100	0	468	100

Figure 1

Self-citations to all volumes in volumes published 1960-70

×——× Journal of Applied Bacteriology
⊙- - -⊙ Journal of Applied Chemistry
+·-·-·+ Journal of the London Mathematical Society



Read as follows: 11% of all articles published in JAB account for 52% of all citations in JAB to articles published in JAB.

Tables 5-7 and Figure 2 give the results of the analysis of citations to the years 1951-59 only. The difference between the distributions for JAC and JLMS is now slight, but there is a very wide difference between them and JAB.

Within each journal, the differences between the overall distribution and that for 1951-59 are not great; in the case of JAB, where citations go back furthest (to 1938) and where one might therefore expect the greatest differences, they are negligible.

Table 5

Journal of Applied Bacteriology:
citations to articles in JAB 1951-59 by articles in JAB 1960-70

	Articles cited			Citations		
No. times cited	N (cum)	N (cum)	% (cum)	N	N (cum)	% (cum)
17	1	1	0.3	17	17	4.2
9	1	2	0.5	9	26	6.4
8	2	4	1.1	16	42	10
7	2	6	1.6	14	56	14
6	3	9	2.5	18	74	18
5	4	13	3.6	20	94	23
4	13	26	7.1	52	146	36
3	18	44	12	54	200	49
2	59	103	28	118	318	78
1	91	194	53	91	409	100
0	172	366	100	0	409	100

Table 6

Journal of Applied Chemistry:
citations to articles in JAC 1951-59 by articles in JAC 1960-70

	Articles cited			Citations		
No. times cited	N	N (cum)	% (cum)	N	N (cum)	% (cum)
6	1	1	0.1	6	6	2.3
5	1	2	0.2	5	11	4.2
4	8	10	1.0	32	43	16
3	11	21	2.1	33	76	29
2	40	61	6.2	80	156	59
1	107	168	17	107	263	100
0	810	978	100	0	263	100

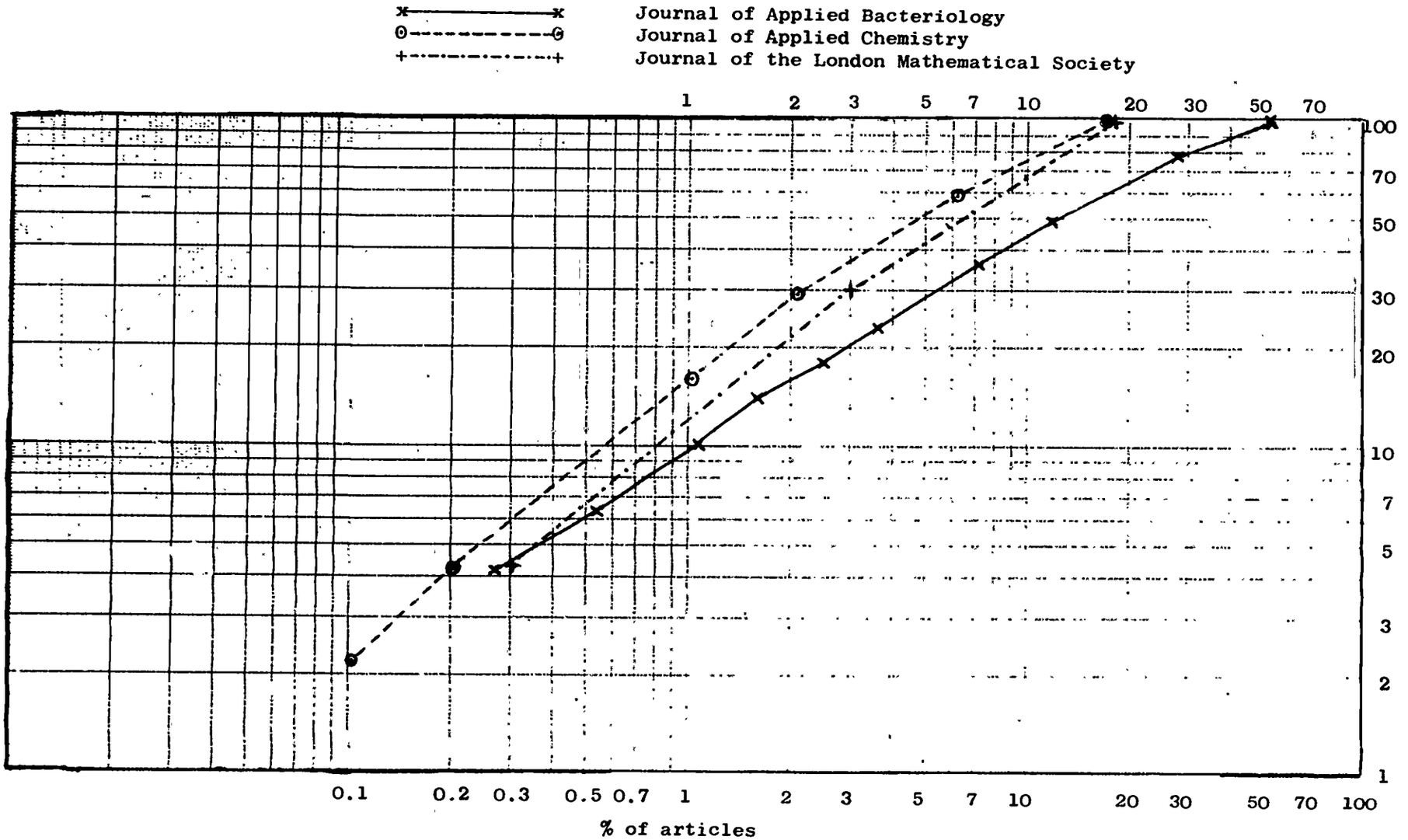
Table 7

Journal of the London Mathematical Society:
citations to articles in JLMS 1951-59 by articles in JLMS 1960-70

	Articles cited			Citations		
No. times cited	N	N (cum)	% (cum)	N	N (cum)	% (cum)
3	2	2	0.3	6	6	4.3
2	18	20	3.0	36	42	30
1	99	119	18	99	141	100
0	552	671	100	0	141	100

Figure 2

Self-citations to 1951-59 volumes in volumes published 1960-70



Read as follows: 12% of all articles published in JAB 1950-59 account for 49% of all citations in JAB to articles in those volumes.

It was possible to carry out a partial test to see how representative the self-citations are of all citations; citations made in 1969 and 1970 to all the articles in the 1960 volumes of all three journals (with the addition of the 1961 and 1962 volumes of JAB) were compared with all the citations to these volumes in Science Citation Index for the same year. The results, given in Table 8, compare interestingly with the Xhignesse & Osgood figures on p. 5, and support the estimate (made, incidentally, before the figures in Table 8 were calculated) that between 2 and 10 per cent of all citations are self-citations. The articles that are self-cited tend to be more productive of citations than the others. This would tend to give a greater concentration on a limited number of articles than appeared from self-citations. In fact, the pattern shown by SCI (see Table 9) was not very different from the pattern shown by self-citations, but the difference in the periods covered would distort the picture.

Table 8

Comparison of self-citations
with citations in Science Citation Index

Source of cited articles	No. of articles			No. of citations		
	(a) cited in SCI 1969-70	(b) cited in 'home' journal 1969-70	(b) as % of (a)	(c) in SCI 1969-70	(d) in 'home' journal 1969-70	(d) as % of (c)
JAB 1960	31	4	13.9	97	4	4.1
JAB 1961	18	4	22.2	48	4	8.3
JAB 1962	30	6	20	76	8	10.5
JAB 1960-62	79	14	17.7	221	16	7.2
JAC 1960	20	3	15	33	3	9.1
JLMS 1960	22	4	18.2	42	5	11.1

Table 9

Pattern of citations as shown by Science Citation Index

Citations to	% of articles	account for	% of citations
JAB 1960-62	3.2		30
	5.3		41
	7.4		50
	10		59
	15		73
	22		84
	41		100
JAC 1960	3.3		39
	8.5		71
	18		100
JLMS 1960	3.1		24
	9.4		52
	16		71
	34		100

From this evidence, it would appear probable that considerable concentration of citations occurs within journals, not to the same degree as between journals, but sufficient, if further study shows similar results, to justify the production of volumes of 'key articles'. A 10 per cent selection might cover between about 50 and 75 per cent of citations, depending on the subject and/or the journal. These results, preliminary though they are, make an interesting comparison with the results of general citation analyses given on p. 1, and would certainly justify a much larger test of the hypothesis.

2.3 Consistency over time

Two further tests were carried out, to see whether the items cited soon after publication were similar to those cited after an interval. This is important, since even if there is considerable concentration of citations on a relatively few articles, this is of little practical use unless citations show consistency over time. Otherwise, a selection made at one point in time of key articles from, say, a decade on the basis of citations would be valid only for that point in time, or a few

years after. Ideally, one would carry out a rank correlation on citations, made at two different points in time, to articles published in a given period; only articles cited three or more times in either period would be included. However, the number of citations in the present study was too small to permit such a test.

First, citations to the 1959 and 1960 volumes of each journal appearing in 1960-62 were compared with those appearing in 1968-70. The results, given in Table 10, are rather surprising. The great majority of articles were cited in only one of the two periods, and a sizable number were cited only in the later period. Some articles that were quite heavily cited in the earlier period were not cited at all in the later (this applies to 7 of the 10 that were cited three times or more).

Table 10

Comparison of citations made in two periods

	Number of 1959-60 articles cited			
	In both periods	In 1960-62 only	In 1968-70 period only	Total
JAB	10	33	12	55
JAC	6	48	5	59
JLMS	3	11	3	17

Secondly, all 1938-59 articles in JAB cited three or more times in JAB 1960-62 were compared with all 1938-59 articles cited in 1968-70. (JAB was chosen because the number of citations was larger than in JAC or JLMS; but it should be remembered that JAB had a lower concentration of citations than the other two journals). Only 16 articles were cited to this extent in the earlier period; of these, only one (the most heavily cited) was cited three or more times in the later period, and 10 were not cited at all. Only 2 articles were cited three or more times in the later period; one of these was, as already mentioned, heavily cited in the earlier period, but the other received only one citation in the earlier period. Results were little more satisfactory when articles

cited twice or more were considered; 34 satisfied this criterion in the earlier period, and 10 in the later period, but only 4 were cited twice or more in both periods.*

Evidently some articles decay, while others mature. These results, based though they are on very small numbers, do cast doubt on the possibility of selecting key articles, at any rate until some years have elapsed. Further study is needed of this important aspect; as Sandison has pointed out[10], the Bradford distribution loses some of its potential practical value if the contents of the Bradford set do not show consistency over time.

---

*These findings are in agreement with those of a much larger study by Cole.[3] He found a correlation coefficient of 0.72 between citations in 1964 and 1966 to articles published in the Physical Review for 1963. He also drew from Science Citation Index a 10 per cent sample of papers (from all fields of science) published before 1961 and receiving 10-20 citations in 1966, and all the papers published before 1961 and receiving 30 or more citations in 1966. Of the 587 papers in this sample, 74 (13 per cent) received 3 or fewer citations in 1961. He also has some interesting comments on the reasons for 'delayed recognition'.

3. OBSOLESCENCE AND GROWTH

The data were collected in such a way that it was possible to examine the obsolescence rates and to look both for differences between the journals and for changes with the passage of time. Although this was not the main point of the exercise, it seemed useful to take the opportunity of using the data to do so, always remembering that they represent only within-journal citations in the journals in question. There are however some advantages in this limitation, for the numbers of items available for citation were precisely known. It is not improbable that obsolescence rates will differ from one discipline to another. Few journals restrict their contents or readership rigidly within a narrow field, and each paper is likely to cite papers from, and to be cited in papers from, a somewhat different range of subjects. But the within-journal citations are likely to be more homogeneous in their disciplinary affiliations than either all the citations in, or all the citations to, any particular journal.

3.1 Problems of technique

Before useful conclusions could be drawn from the available data, certain problems of technique had to be decided.

(i) Citation probabilities

As has been pointed out elsewhere,[7,11] raw citation counts do not provide a useful measure of changes in the interest shown in particular volumes or articles. It is necessary to allow for the bias in the raw data resulting from differences in the numbers of items of each age available for citation. Similar considerations apply to a comparable bias resulting from differences in the numbers of items in which the citations might have been made. Accordingly the citation counts were reduced to citations-per-paper. The count in any one volume to any other volume was divided both by the number of papers in the cited volume and by the number in the citing volume, and then for convenience multiplied by 10^5. This gave an estimate of the probability that any one of the available papers would be cited in any one of the papers examined. These probabilities are directly comparable between any of the cells in Tables 11-13, which give the full matrices of results for the three journals.

Table 11

Journal of Applied Bacteriology: self-citation matrix

Cited vols.	Total no. articles published	No. of citations appearing in each year: 1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970
		Citations per paper per paper x 10^5										
1938	17		2 271	1 120			1 100					
1939	17											
1940	14								1 115		1 108	
1941	27								3 179	1 53		
1942	16					1 112				1 89		
1943	25	1 75				1 71			1 61			
1944	19			1 107		1 94			1 85			
1945	21			1 97			1 81					
1946	27	2 110				1 66		1 56	3 179	2 106	1 56	
1947	19					1 91		1 80	2 170			
1948	11		1 211		1 111	3 187		1 138	1 117	1 130		
1949	27		2 172	1 76	1 45	1 66	1 63	1 56	3 179			
1950	21			1 97				1 72	1 307			
1951	32	1 236	1 73	2 128	1 152	1 56	4 212		2 101	2 89	1 47	1 28
1952	19	6 396		1 107	3 193			2 159	3 255	2 150		
1953	15	3 377	1 155	2 272	1 81	1 119		1 101		2 190		
1954	37	1 201	1 63			1 18	3 137	1 41	2 87			
1955	65	9 261	7 250	5 157	7 131	3 82	7 182		6 149	3 66	1 23	3 42
1956	49	7 270	12 570	10 416	4 100	7 255	10 346	4 124	8 263	7 201	3 92	7 129
1957	61	8 217	13 197	9 301	6 120	1 117	3 83	9 224	9 238	4 94	2 50	15 222
1958	43	12 327	6 324	3 142	5 142	3 125	3 118	2 70	1 150	5 166		1 21
1959	45	18 755	9 465	6 272	7 190	7 278	6 226	6 202	5 179	6 193	1 34	7 140
1960	53	17 605	17 746	14 539	17 391	20 674	10 320	12 343	13 396	15 404	4 114	5 85
1961	43		9 487	11 522	11 312	11 457	5 197	4 141	5 188	2 66	1 35	3 63
1962	49			22 916	11 274	20 729	8 277	15 464	9 296	9 262	5 155	3 55
1963	82				12 178	12 261	11 227	17 314	12 236	6 104	7 129	12 132
1964	56					7 223	18 545	18 487	9 259	8 230	5 135	2 32
1965	59						8 229	19 488	14 383	17 412	9 231	12 183
1966	66							13 298	12 293	21 456	9 206	13 177
1967	62								4 104	16 369	11 269	15 218
1968	70									4 82	10 216	13 167
1969	66										1 23	11 150
1970	111											16 130
Total citations		91	81	90	90	106	99	128	136	134	72	139
Citations per paper citing		1.72	1.88	1.84	1.10	1.89	1.68	1.94	2.19	1.91	1.09	1.25

Table 12

Journal of Applied Chemistry: self-citation matrix

Cited vols.	Total no. articles published	No. of citations appearing in each year / Citations per paper per paper x 10^5										
		1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970
1951	129		1 / 10	1 / 8	1 / 9	2 / 18	2 / 18	1 / 9	1 / 10	2 / 20	1 / 10	
1952	133	7 / 54	2 / 18	3 / 23	1 / 8	1 / 9		2 / 18	4 / 38	4 / 38		
1953	87	6 / 71	6 / 84	4 / 47		2 / 26	3 / 39				2 / 29	
1954	108	3 / 29	5 / 56	4 / 38	1 / 10		1 / 11	4 / 44	4 / 47		1 / 12	
1955	95	3 / 33	2 / 26	3 / 33	1 / 12	1 / 12	2 / 24	1 / 12	3 / 40		2 / 27	
1956	82	3 / 38	4 / 60	3 / 38	1 / 14	1 / 14	2 / 28		1 / 15		1 / 15	2 / 30
1957	101	9 / 92	3 / 36	6 / 61	6 / 66	3 / 34	3 / 34	2 / 23	2 / 25	2 / 25	1 / 13	1 / 12
1958	135	17 / 130	9 / 81	8 / 61	4 / 33	6 / 51	3 / 25	1 / 9	3 / 28	6 / 56	5 / 47	1 / 9
1959	108	23 / 219	14 / 158	13 / 124		1 / 11	1 / 11			1 / 12	4 / 47	2 / 23
1960	97	5 / 53	11 / 138	9 / 96	3 / 34	2 / 24	9 / 105	1 / 12	1 / 13	4 / 52	4 / 52	
1961	82		12 / 179	16 / 201	8 / 108	9 / 126	7 / 97	6 / 86	1 / 15	8 / 124	1 / 15	1 / 15
1962	97			13 / 138	5 / 57	11 / 138	11 / 129	14 / 170	1 / 13	6 / 78	5 / 65	8 / 102
1963	90				3 / 37	18 / 230	13 / 164	3 / 39	10 / 140	14 / 197	8 / 113	1 / 14
1964	87					7 / 92	16 / 209	9 / 122	10 / 145	11 / 160	8 / 116	4 / 57
1965	88						14 / 181	14 / 187	16 / 230	11 / 158	4 / 57	
1966	85							8 / 111	7 / 104	5 / 74	3 / 45	2 / 29
1967	79								4 / 64	20 / 320	5 / 80	4 / 63
1968	79									9 / 144	16 / 256	10 / 156
1969	79										17 / 272	13 / 203
1970	81											15 / 229
Total citations		76	69	83	34	64	87	66	68	103	88	64
Citations per paper citing		0.78	0.84	0.86	0.38	0.74	0.99	0.78	0.86	1.30	1.11	0.79

Table 13

Journal of the London Mathematical Society:
self-citation matrix

Cited vols.	Total no. articles published	No. of citations appearing in each year 1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	Citations per paper per paper x 10^5
1940	44				1 25				1 12				
1941	44			1 27				1 15			2 32		
1942	35												
1943	41			1 30	1 27	1 15		1 17	1 13	1 12			
1944	44				2 51	1 11		2 31			1 16		
1945	39				1 29		1 17		2 28	1 12			
1946	50		1 29	1 24				2 27		1 10	1 14		
1947	65			2 37						1 7	1 11	1 10	
1948	58		3 74	4 84		2 21	1 11		1 9	1 8			
1949	49	1 28				1 12				3 29	2 28		
1950	61	1 22			1 18	1 10	1 11				1 12		
1951	60	2 45				1 10					2 23	1 11	
1952	91	2 30	1 16	2 27			1 7	1 7	1 6		2 15	1 7	
1953	83	1 16		1 15	2 27	2 14	6 47	1 8			2 17		
1954	74	1 18	1 19	2 33		2 16	1 9	2 19	2 15	1 7			
1955	72	2 38	2 40	1 17		1 8	3 27			1 7	2 19	1 9	
1956	77	4 70		3 47	1 14	4 31	2 17		1 7	2 12			
1957	81	5 83	2 35	1 15	1 14		2 16	3 25	2 13			1 8	
1958	79	3 51	1 18	2 31	3 42	1 8	2 17	1 9	5 34	2 12	1 9		
1959	74	5 91	3 58		1 15		3 27	1 9	2 15	1 7		2 17	
1960	74	2 37	1 19	3 49	4 60	4 33	2 18	6 56	2 15	1 7	5 47		
1961	70		3 61	2 35	2 32	2 17	1 9	2 20	2 16	2 14	2 20	7 64	
1962	82			2 30	3 41	10 74	7 56	4 33	3 20	3 18	4 34	1 8	
1963	90				2 25	3 20	3 22	5 38		2 11		1 7	
1964	166					5 16	17 67	8 33	6 20	6 17	2 8	1 4	
1965	153						4 17	6 27	7 25	4 13	5 23	5 21	
1966	146							4 19	6 22	12 39	5 24	1 4	
1967	184								8 24	3 8	6 23	3 10	
1968	209									0 0	11 37	4 12	
1969	143										7 34	10 45	
1970	157											3 12	
Total citations		29	18	28	25	41	57	50	52	48	65	43	
Citations per paper citing		0.39	0.26	0.34	0.28	0.25	0.37	0.34	0.28	0.23	0.45	0.27	

(ii) Exponential decay

It is generally maintained that the frequency of citation decays exponentially with increasing age and that citation probabilities behave similarly. Exponential curves become linear if the logarithms of the observed values are plotted against time (e.g. on semi-log paper) and the probability decay rate (obsolescence rate) is then measured by the slope (regression constant) of that line. Access to a pre-programmed calculating machine made the computation of regression constants relatively easy: their antilogs are the exponential decay factors.

(iii) Treatment of zero counts

Tables 11-13 contain a considerable number of cells with zero values. These are not susceptible to logarithmic transformation, because the log of 0 is minus infinity which can be neither plotted on a graph nor used in calculations. A usual procedure[13] for eliminating zeros is to add a small arbitrary figure to all the observations, but this is inappropriate when the zeros are numerous and concentrated in one section of the data, because in these circumstances the particular arbitrary value selected materially influences the size of the regression constant. A frequent alternative procedure is to combine adjacent cells of the table and use the means of their individual values. This is legitimate if the differences between the cell characteristics are additive, but this is not true of points on an exponential curve before they are transformed to logs. The discrepancies cancel out if the whole table is condensed by combining equal numbers of cells throughout. But to eliminate all zeros in the present data would mean reducing each table to only two or three age groups from which no regression constants of value could be obtained. Nor is it appropriate just to ignore the data from the older literature where most of the zeros occur. It is, however, possible[12] to plot the 'appropriate value' for the observations in the combined cell against a weighted* mean of the ages, and this was done. For citation probabilities the 'appropriate value' for combining cells cannot be obtained simply by taking the mean of the

---

*Representing the age corresponding to the point on the exponential curve at which the arithmetic mean of the appropriate points on the theoretical curve would fall. It is necessary to have an approximate value of the decay rate in order to calculate the weighted means.

separate probabilities, but must be recalculated from the citation counts divided by the total numbers of citing papers and of papers available for citation for the whole of the combined cell. One other point needs to be remembered, that any combination of cells reduces the variance of the resulting parameters.

3.2 Diachrony and synchrony

The results of this study can be interpreted only in the light of a clear understanding of the possible influences of time on citation frequencies. Two different effects must be clearly distinguished. First, in any one list of references, the probability that any one of all the potentially available papers will be cited is usually lower the greater its age. This is the synchronous situation observable in data collected from a single source. It indicates the relative importance to the author(s) at that time of literature of different ages. Secondly, if the citations to any one item are followed through successive lists of references, they tend to get fewer as time passes. This, the diachronous situation, can be examined only by comparing sources published over a period of years. The diachronous situation is subject to more complex influences than the synchronous one, and has as yet hardly been studied. Nonetheless there has been a tendency to assume that synchronous data can be used to forecast the diachronous situation in library planning.

The complexity of the interrelation between synchronous and diachronous decay of citation probabilities can best be explained through the examples in Table 14, set out similarly to Tables 11-13, in which columns represent synchronous, and rows represent diachronous, sets of data.

Table 14

Interrelation of synchronous and diachronous decay

	A					B					C				
Year	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1	30	20	13	8½	5½	30	12	8	8½	10	22	16	11½	8½	6
2		30	20	13	8½		30	17	13	13		24	18	13	9½
3			30	20	13			30	20	17			27	20	14½
4				30	20				30	23				30	22
5					30					30					33

In section A the vertical synchronous exponential decay rates are identical in each year: as will be seen the horizontal diachronous figures correspond exactly. In section B the synchronous exponential decay rates were falling year by year: the diachronous figures no longer correspond, and their changes are neither uniform nor exponential. In section C another variable has been introduced, by increasing steadily from year to year the total number of citations per recent paper, but maintaining a constant decay rate, in each column. The diachronous decay rate remains exponential, but is different from the synchronous one. It might be expected that citations-per-paper would increase as the total number of citable articles increases: if they are held down artificially by editorial policy this might reduce the references to older papers (which can be traced through the citations in younger ones) and therefore increase synchronous decay rates.

On general grounds therefore there are no inherent reasons why synchronous and diachronous decay rates should be identical and several why they might not. It is to be expected too, that since diachronous rates are influenced by more variables, their variance would be higher. and the confidence limits wider, than for synchronous figures.

3.3 Results

The present data were examined to see if there was any evidence for changes in synchronous decay rates with time, for the relation between diachronous and synchronous rates, to see whether synchronous rates were uniformly exponential, and for any differences between the three journals, as well as for changes in the numbers of papers per year and citations per paper.

(i) Growth of the literature

It is interesting to note the changes in numbers of articles per year in each journal. Table 15 summarizes the picture (using overlapping periods to avoid distortions due to freak years).

Table 15

Average number of articles per year

	JAB	JAC	JLMS
1941-50	21	-	49
1946-55	27	-	66
1951-60	42	108	77
1956-65	54	97	95
1961-70	66	85	140

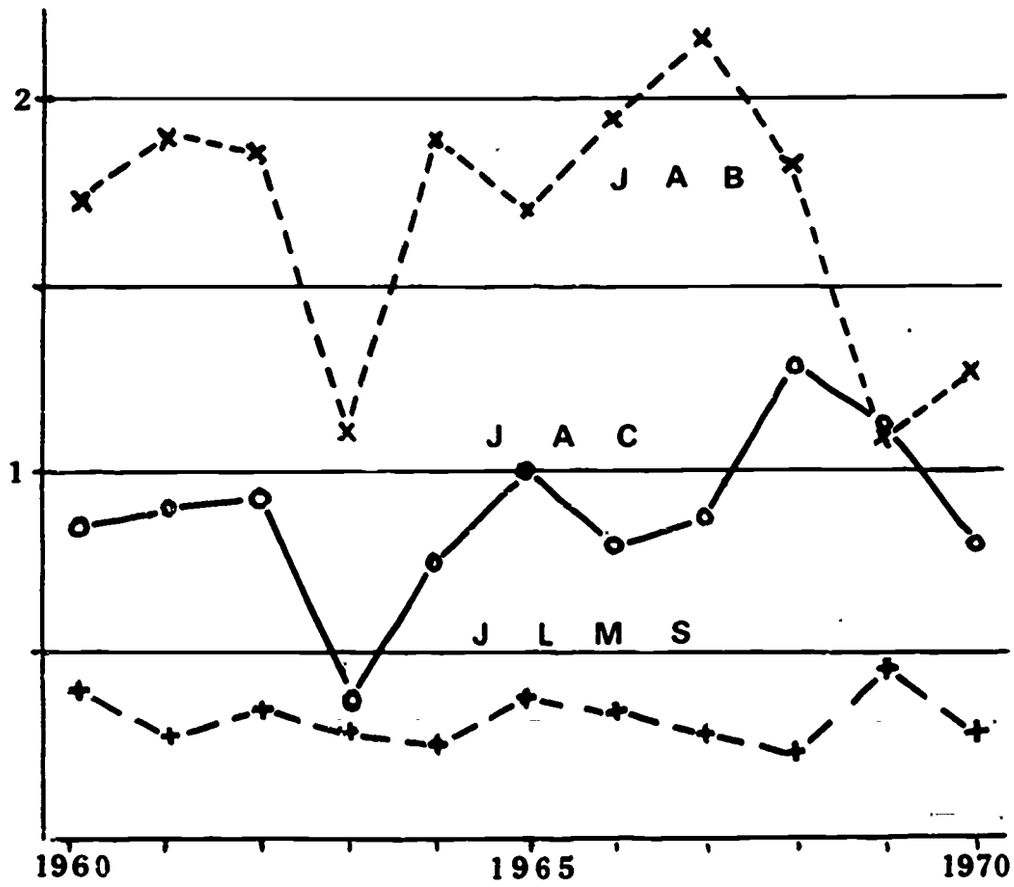
It will be noted that while JAB and JLMS grew quite fast (at roughly similar rates), JAC actually got smaller.

(ii) Citations per paper

As Figure 3 shows, there was no evidence for any consistent tendency for the within-journal citations per paper to alter in any of the three journals. The mean values for the journals were significantly different.

Figure 3

Within-journal citations per paper



(iii) Changes in synchronous decay rates with time

Separate annual decay rates for within-journal synchronous citation probabilities were calculated for each year for each journal, using all the available data. There was no evidence for any tendency for those rates either to increase or to decrease over the 11 years 1960-1970, in any of the journals, whether calculated from all the data or for the shorter periods of ages 1-6, 1-12, 1-18.

(iv) Diachronous decay

There was relatively little data for direct estimation of within-journal diachronous citation probability decay rates. The blocks of data represented by ages 1-6, cited in 1963-1968 or published in 1959-64, had the maximum number of common records for direct comparison of synchronous and diachronous data. The data proved insufficient for any differences between the synchronous and diachronous rates within any of the journals to be significant.

(v) Uniformity of exponential decay rates

Table 16

Within-journal annual citation probability synchronous decay factors and 5 per cent confidence limits

	JAB	JAC	JLMS
Full age range	0.895 (0.91 - 0.88) n = 179	0.83 (0.885 - 0.78) n = 132	0.935 (0.95 - 0.92) n = 184
Ages 1-6	0.855 (1.07 - 0.68) n = 66	0.72 (0.805 - 0.655) n = 64	0.86 (0.96 - 0.775) n = 61
Ages 7-12	0.88 (0.985 - 0.78) n = 55	0.86 (0.88 - 0.84) n = 52	0.88 (0.99 - 0.78) n = 46
Ages 13-18	1.055 (1.31 - 0.85) n = 35	0.91 (1.08 - 0.765) n = 16	0.91 (1.10 - 0.74) n = 28
Ages 19+	0.955 (1.10 - 0.83) n = 25		0.93 (1.00 - 0.835) n = 31

Factor of 1 = no decay or growth. Factor increases with decreasing decay rate.

Table 16 sets out the synchronous within-journal citation-decay rates and their 5 per cent confidence limits, both for the full data and for 6 year spans of increasing age. Figure 4 shows in graphical form the citation probabilities after combining the 11 cells for each age. In the absence of any trend in the decay rates for separate years, this diagonal summation of cells does not require corrections for exponential changes, but the variance is substantially reduced. Differences therefore reach the 5 per cent significance levels between the age groups in this combined data which do not do so in the analysis of the full data. Even in the full data, the decay rate for the Journal of Applied Chemistry is significantly greater for ages 1-6 than for ages 7-12. It is notable too that for each journal the decay rate falls (with one exception) as the ages increase. On a sign test, this tendency is almost significant at $P < .05$ from the eight paired comparisons. This was examined in another way, by calculating, from the data summed over the 11 years of citation, regression constants as each additional year of age was brought into account. For the Journal of Applied Chemistry, on bringing

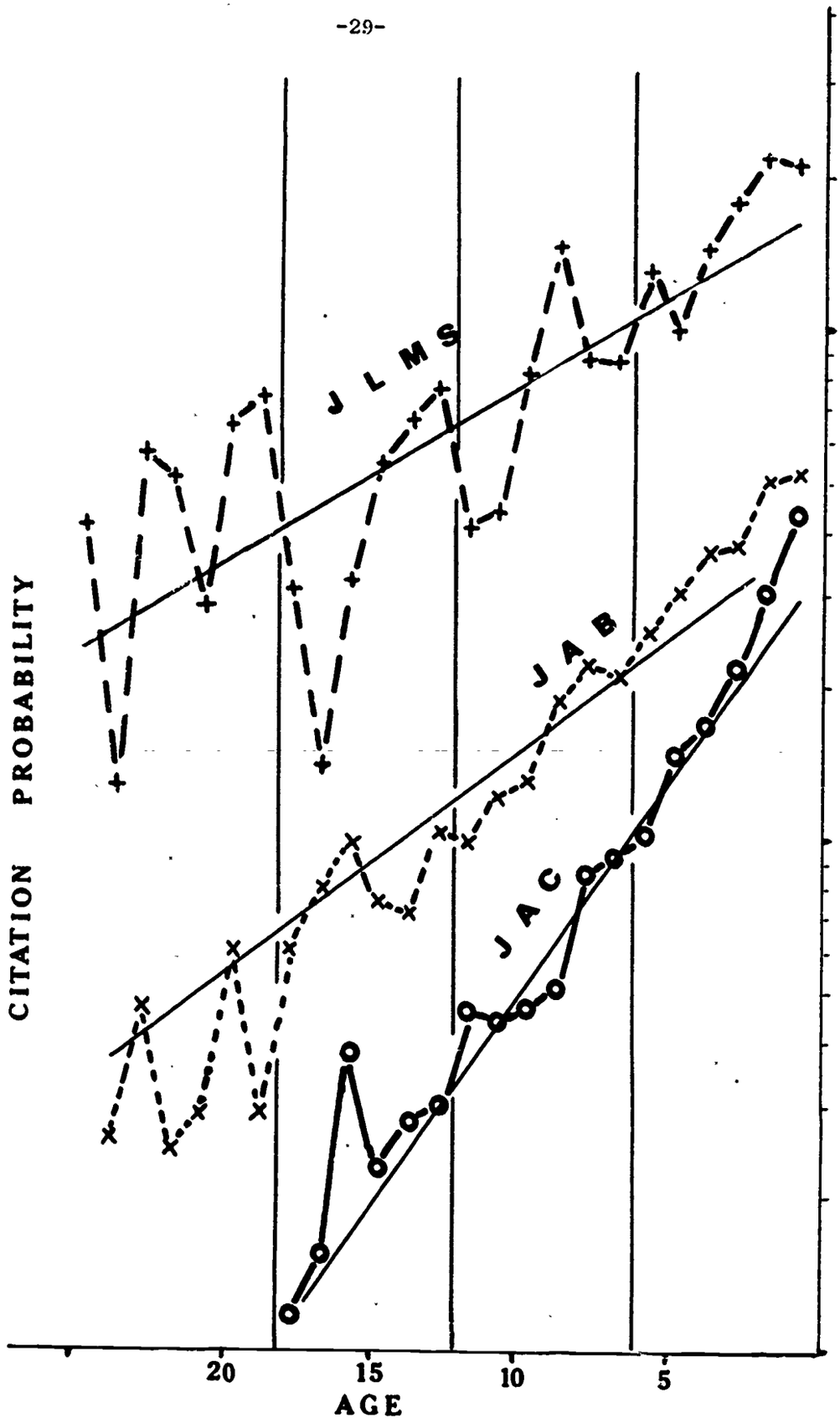


Figure 4

Age distribution of within-journal citations
(log probabilities by age)

in additional years of increasing age, the regression constant fell steadily from ages 3-18: bringing in years of reducing age, it fell from ages 15-9 and then rose steadily to age 1. The Journal of the London Mathematical Society showed almost continuous increases in the regression constant as additional years of reducing age were brought in from age 16 to age 1: bringing in years of increasing age showed wider fluctuations, but a general downward trend. For the Journal of Applied Bacteriology the pattern was similar but the actual changes in the values of the regression constants much smaller.

It seems necessary to conclude therefore that for all three journals the probability that any particular paper will have been cited falls most rapidly with increasing age for articles only a few years old, and that annual decay in the citation probability gets less with increasing age. The decay rates are not therefore strictly exponential. It must be emphasized that this data relates only to within-journal citations. It will be interesting to see whether similar conclusions apply to citations to other journals - but experimentally there are considerable difficulties in obtaining reliable estimates of the numbers of papers available for citation, without which citation probabilities cannot be calculated.

It will be noted that zero decay (i.e. decay factor = 1.0) is within the confidence limits for all three journals for citations to items over 12 years old.

(vi) Differences between journals

Over the full age ranges the annual citation-probability synchronous decay rates for the three journals are significantly different, but this is almost entirely due to differences affecting the most recent literature. The more rapid decay in the use of recent literature, which has been described as an immediacy effect[9], was most marked in the JAC: for the full age range its decay rate might have been less if it had been in publication longer and there had been articles over 19 years old available for citation. The decay rates for the JLMS for the separate 6 year groups are probably too great by an accident of grouping, as can be seen from Figure 4.

It may help some readers to visualize the situation to say that over ages 1-6 the probability of citation fell to half in 2¼ years in the Journal of Applied Chemistry but in 4½ years in the other two journals. Over ages 7-12 years it fell to half in about 6 years in all three.

4. CONCLUSION

The results of this small study show that some of the so-called 'laws' of information science need to be interpreted with care and discrimination.

The hypothesis that cited articles are quite heavily concentrated within as well as between journals, and that that concentration extends to individual articles as well as to volumes, receives some support, and thereby apparently extends the application of 'Bradford's Law'. But it seems that it may be difficult to identify key articles until some years after publication and those that are most significant at one time are not necessarily the same as those some years later. The idea of publishing collections of key articles from journals deserves further study; in particular, differential use over time needs to be explored. This could relatively easily be done from the use of Science Citation Index tapes, provided that some suitable estimate of the numbers of papers available to do the citing can be obtained.

The decay in the frequency of citations to volumes with age has again been demonstrated, but it does not follow an exact exponential law. Rates measured from recent volumes are shown to underestimate the importance of older ones, and rates for one discipline are inappropriate to another. Furthermore there appears to be no inherent reason why an obsolescence rate measured at any one time should reflect future changes in the citation of particular volumes closely, although in the limited data available in this study it appeared to do so.

REFERENCES

[1]BURTON, Robert E. Citations in American engineering journals, I: chemical engineering; II: mechanical engineering; III: metallurgical engineering. American Documentation, 10(1), January 1959, 70-73; 10(2), April 1959, 135-137; 10(3), July 1959, 209-213.

[2]COILE, Russell C. Information sources for electrical and electronics engineers. IEEE Transactions on Engineering Writing and Speech, EWS 12(3), October 1969, 71-78.

[3]COLE, Stephen. Professional standing and the reception of scientific discoveries. American Journal of Sociology, 76(2), September 1970, 286-306.

[4]CRAIG, J.E.G. Characteristics of use of geology literature. College and Research Libraries, 30(3), May 1969, 230-236.

[5]KESSLER, M.M. and HEART, F.E. Analysis of bibliographic sources in the Physical Review. Cambridge, Mass., MIT, Lincoln Laboratory (Report R-3, July 13, 1962).

[6]KOURILOVA, Magda. Analyze pramenov vedeckych informacii. Metodika a Technika Informaci, 1-2, 1965, 35-70.

[7]LINE, M.B. The 'half-life' of periodical literature: apparent and real obsolescence. Journal of Documentation, 26(1), March 1970, 46-54.

[8]MARTYN, John and GILCHRIST, Alan. An evaluation of British scientific journals. (Aslib Occasional Publication no.1). London, Aslib, 1968.

[9]PRICE, D.J. de S. Networks of scientific papers. Science, 149(3683), July 1965, 510-515.

[10]SANDISON, A. Library optimum (letter). Nature, 234(5328), December 10 1971, 368-369.

[11]SANDISON, A. The use of older literature and its obsolescence. Journal of Documentation, 27(3), September 1971, 184-199.

[12]SANDISON, A. Treatment of zero counts in obsolescence studies. Unpublished.

[13]SOKAT, R.R. and ROHLF, F.J. Biometry. San Francisco, Freeman & Co., 1969, p.384.

[14]VICKERY, B.C. Indicators of the use of periodicals. Journal of Librarianship, 1(3), July 1969, 170-182.

[15]XHIGNESSE, Louis V. and OSGOOD, Charles E. Bibliographical citation characteristics of the psychological journal network in 1950 and in 1960. American Psychologist, 22(9), September 1967, 778-791.